Contribution of Schools to Covid-19 Pandemic: Evidence from the Czech Republic

Martin Šmíd, Jakub Drbohlav, Milan Zajíček*

July 6, 2021

Main findings

- In-person education is a significant source of infections in the children age cohorts, yet a majority of those infections come for other sources than schools
- The effect increases with the age and is significantly higher in secondary schools
- Data suggest a negligible or zero effect of children attendance to kindergartens
- The difference between opening regimes increase with age
- Data do not support the hypothesis that closing schools causes additional infections in other environments
- A simple indicator of safe opening of schools may be constructed. Using it, we can conclude that closing primary and secondary schools in Czechia during the school year 20/21 was mostly reasonable given the state of surrounding epidemics; however, the spring opening of primary schools could have been quicker

Introduction

Closing schools during the present pandemics is the one of the most controversial issues. According to epidemiological mainstream, schools are strong drivers of respiratory diseases [Gold, 2021, Stein-Zamir et al., 2020, Torres et al., 2020]; see also [Forbes et al., 2021, SAGE, 2020, Ismail et al., 2021, Lessler et al., 2021, Brauner et al., 2021b, Haug et al., 2020], but also [Zhu et al., 2020, Zimmerman et al., 2021];

*Institute of Information Theory and Automation of the CAS, BISOP

therefore, closing schools was one of the first governmental reactions to the outburst of the present pandemic. However, the price the society pays for closing schools is high. Online education is not an equivalent substitute for the in-person one [Di Pietro et al., 2020, Engzell et al., 2020, Maldonado and De Witte, 2020]; moreover, the isolation of students leaves deficits in their necessary social contacts [Bignardi et al., 2020, ECDC, 2020, Di Pietro et al., 2020, Ravens-Sieberer et al., 2021]. Thus, for the society, decision whether and when to close schools means a painful trade-off, making any decision political rather then scientific; science, however, has irreplaceable role in giving the best possible (quantitative) basis for these decisions.

Unfortunately, it is difficult to evaluate the effect of closing schools in the context of the current pandemic. The virus was completely unexplored as it came and its knowledge increased only gradually, so we only gradually get to know whether and how differently the virus affects children in comparison with the rest of the population. To estimate the effect from running epidemic data is also problematic, mainly because the counter-epidemic measures are usually being introduced (and released) together, so it is difficult to distinguish their effects, and even in cases when the measures are applied in different times, still they can be assessed only in the context of the other measures applied. [Brauner et al., 2021a], for instance, regards school closure as a very effective means of curbing the present pandemics; however, it is necessary to take into account that school closures were usually the first measures applied, so their measured effect may be overestimated in comparison with the other measures.

In our analysis, we exploit a "gap" in this puzzle: the fact that certain age cohorts of children nearly uniquely correspond to the degrees of schools and, except for kindergartens, the vast majority of children do attend their classes. Thus it could be expected that closing (opening) of particular degrees will result in significant changes in infections in the corresponding cohorts. The other measures, on the other hand, are usually much less age specific (wearing masks, bans of gatherings, for instance) or are affecting different age cohorts (home office) so they can be expected not to obscure much the effects of school closures.

To evaluate the effects of the closures, we used a simple epidemiological-like model, in which the infections in the children cohorts depend on the overall magnitude of epidemic, its magnitude within the district and within the corresponding cohort, the latter in dependence on the current school opening regime. Using our model, we estimate quantitative impact of various regimes of school opening, and we construct a simple estimator of the infection grow within the cohort, which can be useful for for decisions on school restriction degree during the epidemic. Applying it retrospectively to the school year 2020/21 we conclude, that, yet the school closures surely prevented a higher growth in the age cohort of students, the contribution of school opening in safe regimes contributes much less to the growth then the rest of the epidemics.

We estimate parameters our model by means of the data from the Czech Republic. Here, all the primary and the secondary schools had been closed on March 12, 2020, shortly after the first cases of covid infection appeared.

The schools were opened in May, but only in a very limited regime until the summer vacation. After the vacation, all the schools had been opened for two weeks nearly without any restriction. Next, after the overall infection numbers started to rise, wearing masks in classrooms had been imposed. In the end of October, after the serious increase of infections, all the schools except for kindergartens had been closed. Following a partial amelioration of the epidemic, schools started to be gradually opened in the end of November; during the last weeks before Christmas, the first degree, the last year of the second degree and the last year of the secondary schools had been fully opened with masks, while the remaining years of the second degree (6th to 8th) had been opened in weekly rotation regime. After the Christmas holiday, however, all schools had been closed again except for the first two years of the first degree and the kindergartens, which all remained open until February 27, when also kindergartens have been closed until April 12th, 2021, when the their last year had been open. Moreover, on April 12th, the first degree has been opened in the rotation regime with regular testing in schools. In the following weeks, all the schools had been gradually opened, the second degree starting with the rotation regime. In the second half of May, full regime had been introduced in all schools, starting with masks, and later without masks in classroom, all with regular testing at schools except for kindergartens. See Table 1 for the detailed schedule.

Methods

To determine the influence of attending schools of a given kind, we examine the number of reported infections $X_{i,t}$ in the corresponding age cohort for each district i at time t. Generally, we assume that this number depends on the previous-week number of infections, the contact restriction reported two weeks earlier, and current rate of infectiousness. In addition to the total number of infections in population, we take into account the number of infections in the same district, and the number of infections in the same cohort within the district. See Appendix A for the justification of this general model. We explicitly model the influence of school opening regime on the infection number in the cohort. In particular, we assume

$$X_{i,t} = \alpha_i D_t Y_{t-1} + \beta_i D_t Y_{i,t-1} + \gamma D_t X_{i,t-1} + S_t X_{i,t-1} + e_{i,t} \tag{1}$$

$$D_t = d_t C_{t-2}, \qquad S_t = D_t(\nu N_{t-1} + \mu M_{t-1} + \varrho R_{t-1} + \nu^\star N^\star_{t-1} + \mu^\star M^\star_{t-1} + \varrho^\star R^\star_{t-1})$$

for each district i and time t. Here, Y_t is the number of overall infections within the population at t, $Y_{i,t}$ is the number of infections within the i-th district at t, N_t, M_t, R_t evaluate the degree of school opening without masks, with masks, with weekly rotations (with masks on), respectively (zero means all classes closed, one means all classes opened), $N^\star_t$, $M^\star_t$, $R^\star_t$ stand for the same regimes with the addition that students are regularly tested at schools, $e_{i,t}$ is the

Date	Kindergartens	First degree	Second degree	Secondary
Mar-13	voluntary	closed		
May-04				
May-11			last year max. 15 voluntary	
May-25		max. 15 voluntary		
Jun-08				
Jun-26	end of the school year			
Sep-01	start of the school year, opened			
Sep-17		masks ordered in classrooms		
Oct-05				some closed
Oct-14		closed		
Nov-18		1^{st} 2^{nd} open		
Nov-25				last year open
Nov-30		open	rotations, last yr. open	
Jan-04		1^{st} 2^{nd} open	closed	closed
Feb-27	closed			
Apr-12	last years	rotations		
Apr-26	selected open			
May-03			selected rotations	
May-10	open		rotations	
May-17		open with testing	selected open w. t.	
May-24			open with testing	open with testing
Jun-08		masks not required in all but 3 regions		
Jun-15		masks not required		
June-30	end of school year			

Table 1: Schedule of counter-epidemic measures at schools in 3/2020-6/2021.

error term, and $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \gamma, \nu, \dots, \varrho^\star$ are estimated parameters. Finally, d_t is the (predicted) rate of infectiousness and C_t is the contact restriction at t, see Appendix A for details.

We estimate the impact of school opening four ways: by comparing two sub-cohorts of the examined cohort each following (1) either with the same or with different school regime coefficients $\nu, \mu, \dots, \varrho^\star$, by estimating the parameters from (1) applied to the whole cohort and, finally, by estimating the parameters from simplified version of (1) with homogeneous impact of districts.

Comparison of two sub-cohorts with different regime parameters (later abbreviated as S). Here we consider two sub-cohorts of comparable size, the j-th one following its own version of (1):

$$X^j_{i,t} = \alpha^j_i D_t Y_{t-1} + \beta^j_i D_t Y_{i,t-1} + \gamma^j c_j D_t X_{i,t-1} + S^j_t c_j X_{i,t-1} + e^j_{i,t}$$

$$S^j_t = d_t[\nu^j N^j_{t-1} + \mu^j M^j_{t-1} + \varrho^j R^j_{t-1} + \mu^{\star j} M^{\star j}_{t-1} + \nu^{\star j} N^{\star j}_{t-1} + \varrho^{\star j} R^{\star j}_{t-1}],$$

where c_j is the relative size of the j-th sub-cohort with respect to whole cohort. By subtracting, we get

$$\frac{X^1_{i,t}}{c_1} - \frac{X^2_{i,t}}{c_2} = \alpha'_i D_t Y_{t-1} + \beta'_i D_t Y_{i,t-1} + \gamma' D_t X_{i,t-1} + U_t X_{i,t-1} + e'_{i,j,t}, \qquad U_t = S^1_t - S^2_t. \tag{2}$$

Clearly, for any pair of the regime coefficients out of $(\nu^1, \nu^2), \dots, (\varrho^{\star 1}, \varrho^{\star 2})$ to be estimable, the corresponding covariate pair $(N^1, N^2), \dots, (R^{\star 1}, R^{\star 2})$ has to differ sufficiently. In the case or kindergartens, for instance, the only regimes having been applied were opening without masks ($N^1_t = N^2_t = 1$), opening only for the last years ($N^1_t = 1, N^2_t = 0$) and closure ($N^1_t = N^2_t = 0$), which is enough for ν^1 and ν^2 to be estimable. Yet only a minority of coefficients happened to be estimable this way, the clear advantage here is that the regime covariates and coefficients which do not differ cancel out, so our estimate is independent both of these parameters and on the choice of corresponding regime covariates, which are set subjectively sometimes, for instance when the only a subset of classes is opened. Moreover, even though we do not explicitly assume an identical impact of the external environment on both the sub-cohorts, these impacts are likely to be are similar, so they likely more or less compensate by the subtractions, disturbing less the impact of the regime terms. Note that the results of the estimation can also be used to test the null hypothesis that the corresponding regime has no influence on infections, which will be rejected by significant values of either ν^1 or ν^2.[1]

Comparison of two sub-cohorts with common regime parameters (SC). If we decrease the generality of (2) by assuming that $\nu^1 = \nu^2 = \nu, \dots, \varrho^{\star 1} = \varrho^{\star 2} = \varrho^\star$, we get

$$U_t = d_t[\nu \Delta N_{t-1} + \dots + \varrho^\star \Delta R^\star_{t-1}] \qquad \Delta N_t = N^1_t - N^2_t, \dots, \Delta R^\star_t = R^{\star 1}_t - R^{\star 2}_t$$

from which we can estimate those of the regime coefficients for which the corresponding covariates differ at least for some time. Again, the significant values

[1] The significance level has to be corrected here.

may serve for the proof that the opening in the particular regime influences infections.

Estimation from (1) for the whole cohort (W). For the coefficients corresponding to the regimes which had never differed between the sub-cohorts, there is nothing to do but to estimate them directly from (1).

Estimation from a homogeneous version of (1) (WH). To get global results rather than the district-level ones, we decrease the generality of (1) and assume a homogeneous model:

$$X_{i,t} = \alpha h s_i D_t Y_{t-1} + \beta h D_t Y_{i,t-1} + \gamma D_t X_{i,t-1} + S_t X_{i,t-1} + \epsilon_{i,t}. \tag{3}$$

Here, h is the relative size of the examined cohort with respect to the rest of the population and s_i is the relative size of the i-th district's population. By summing (3) over i, we get

$$X_t = \omega h D_t Y_{t-1} + (\gamma D_t + S_t) X_{t-1} + \epsilon_t, \qquad \omega = \alpha + \beta, \tag{4}$$

where X_t is number the overall cohort infections, which may be alternatively interpreted as

$$\begin{aligned} &\mathbb{P}[\text{infection in cohort at } t] \\ &\doteq \omega D_t \mathbb{P}[\text{overall infection at } t-1] + (\gamma D_t + S_t)\mathbb{P}[\text{infection in cohort at } t-1] \end{aligned} \tag{5}$$

(to see it, divide (4) by the cohort size). Similarly, we can interpret (3) as

$$\begin{aligned} &\mathbb{P}[\text{infection in the cohort in district } i \text{ at } t] \\ &\doteq \alpha s_i D_t \mathbb{P}[\text{overall infection at } t-1] + \beta D_t \mathbb{P}[\text{infection in district } i \text{ at } t-1] \\ &\qquad + (\gamma D_t + S_t)\mathbb{P}[\text{infection in the cohort in } i \text{ at } t-1] \end{aligned} \tag{6}$$

(again, it suffices to divide (3) by the cohort size). It is clear from (5) and (6) that quantities αD_t, βD_t, γD_t, $\nu d_t, \dots, \varrho^\star d_t$ may be seen as weights, contributing to the infection in the cohort in the hypothetical case that the probability of infection is the same in the population, in the district and in the cohort.

Finally, if we divide (4) by X_{t-1}, we get the expected relative growth of infections within the cohort:

$$\rho_t \stackrel{\text{def}}{=} \mathbb{E}\left(\frac{X_t}{X_{t-1}} \middle| \text{history up to } t-1 \right) = G_t + S_t \qquad G_t = D_t \left[\omega \frac{h Y_{t-1}}{X_{t-1}} + \gamma \right], \tag{7}$$

from which we can distinguish the part of the growth caused by school opening (quantified by S_t) from that originated outside schools (G_t). In the special case that no contact restriction is applied ($D_t = d_t$) and that the ratio of the infected within the cohort is the same in as that within the whole population (i.e. $\frac{hY_{t-1}}{X_{t-1}} = 1$) we get

$$\rho_t = d_t(\omega + \gamma + \nu N_{t-1} + \dots + \varrho^\star R^\star_{t-1}) \tag{8}$$

i.e. the regime coefficients themselves may serve as weights.

Data

In the Czech Republic, numbers of children attending particular schools degrees correspond well with particular age cohorts. Each September first, children who completed their sixth year are obliged to enter the first grade of an elementary school. When the parents wish, a child who is six as late as by the end of the year can enter the first class, too, some children, on the other hand, start their school later for various, mostly developmental, reasons. If we neglect these exceptions, we can conclude that the first year pupils are six or seven; consequently the vast majority of the first degree (first five grades) pupils belong to the age cohort $6-11$. Consequently, the vast majority of the second degree (the sixth to the ninth year) students belong to the cohort $11-15$, and the students of secondary schools, which typically take four years in the Czech Republic, belong to the cohort $15-19$. As for the kindergartens, where the usual lowest age for admission is three, the vast majority of children attending kindergartens falls into the cohort $3-6$.

For simplicity we assume that the frontier one-year cohorts split by half between the competing school categories, so the infections happening in the cohort split by two between each category. Consequently, the number of cases by pre-school children over week t in the i-th district will be $X_{i,t}^k = Z_{i,t}^3 + Z_{i,t}^4 + Z_{i,t}^5 + \frac{1}{2}Z_{i,t}^6$, the number by the first degree $X_{i,t}^f = \frac{1}{2}Z_{i,t}^6 + Z_{i,t}^7 + \cdots + \frac{1}{2}Z_{i,t}^{11}$, by the second degree $X_{i,t}^s = \frac{1}{2}Z_{i,t}^{11} + Z_{i,t}^{12} + \cdots + \frac{1}{2}Z_{i,t}^{15}$ and that by the secondary schools $X_{i,t}^e = \frac{1}{2}Z_{i,t}^{15} + Z_{i,t}^{16} + \cdots + \frac{1}{2}Z_{i,t}^{19}$,where $Z_{i,t}^j$ is the number of cases by the j-year old individuals in the i-th district. We take the values $Z_{i,t}^j$ from the anonymized person-level list of reported infections [Chech Ministry of Health, 2021] including, among other things, the date of reporting the infection and the age; the series Y_t and $Y_{i,t}$ are computed by their aggregation.

Yet the first cases in Czechia appeared in the beginning of March, 2020, we do not take data until April 5th, 2020 into account, as we regard them as noisy, unreliable and suffering from small sample properties. The last values used for the estimation correspond to the week staring on June 28th, 2021. As we do not use data from 9 weeks of summer vacation and those from the one-week Christmas vacation, this means 53 weeks. As there is 77 districts (including the capital), this gives 4081 observations, from which, however, some additional ones were excluded, as noted below.

The regime covariates have been determined according to measures schedule, listed in Table 1. The values of $N, \dots, R^\star$, which correspond to the full cohorts, were set to 1 if the corresponding regime applied fully, and to 0 if it was not applied at all given week. Some values were set to fractional values; however, some of these observations were excluded from estimation of (1) and (2), especially in the cases when the fractional value could not be determined objectively. We also excluded observations corresponding to holidays and the weeks with no more than two possible school days, as our study concerns only weeks in which the teaching takes place, perhaps online. The list of values of the regime covariates can be seen in Appendix B together with the notes on

	Kindergartens	First degree	Second degree	Secondary
Group 1	last year	1st and 2nd yr.	9th year	4th year
Sub-cohort 1	$\frac{1}{2}(Z^5 + Z^6)$	$\frac{1}{2}Z^6 + Z^7 + \frac{1}{2}Z^8$	$\frac{1}{2}(Z^{14} + Z^{15})$	$\frac{1}{2}(Z^{19} + Z^{20})$
Group 2	5 year old	4th and 5th yr.	7th year	2nd year
Sub-cohort 2	Z^4	$\frac{1}{2}Z^9 + Z^{10} + \frac{1}{2}Z^{11}$	$\frac{1}{2}(Z^{12} + Z^{13})$	$\frac{1}{2}(Z^{17} + Z^{18})$
Period	4/12 – 4/25/21	1/4 – 2/16/21	11/30 – 12/21/20	11/25 – 12/21/20
Full weeks	2	8	3	3
Diff. pars.	ν^1, ν^2	μ^1, μ^2	μ^1, ρ^2	μ^1, μ^2
Common pars.		μ	$\mu - \rho$	μ

Table 2: Details on sub-cohort comparisons

how we determined fractional values and reasons of observation exclusions.

We identified four opportunities for sub-cohort comparison. The first were two weeks starting from April 12th when the kindergartens have been open only for the last (pre-school) year, the second when only the first two years out of the first degree were open in January and February 2021, the third when the second degrees were open for the last year while the rotation regime was applied in the end of 2020, the fourth when only the last year of secondary schools were open from November 25 until the Christmas vacation. As control groups, we chose the closest year (or years) for which their corresponding sub-cohorts do not share a border cohort. For kindergartens, where the attending percentage decreases with the age, we chose four year old as the control sub-cohort, because it is closer to the treatment group. The details on the sub-cohort models are summarized in Table 2. For the values of the covariates, see Appendix B.

Results

We estimated all the S, SC, W and WH variants of our model by weighted least squares where we took $w_{i,t} = \max(20^2, Y_{i,t-1}^2)$ as weights of residual variances. This choice was motivated by residual analysis and our desire to suppress observations suffering from the small-sample property. The detailed results of the estimation may be found in Table 8 (Appendix) and seen in Figure 1 (recall that the regime coefficients always evaluate the rate of infections caused by the own cohort up to d_t, so they are comparable each with other as well as between the cohorts). We see in the Table 8 that the vast majority of regime coefficients comes out positive significant for primary and secondary schools, while the only coefficient ν comes out zero or even significantly negative for kindergartens. We thus conclude that, while we cannot prove the impact of kindergartens, the impact of the primary and the secondary schools appears to be significant.

Further, for all the first degree, the second degree, and the secondary schools, we computed aggregated estimates of the regime parameters as the weighted average of the estimates by the individual models, where we gave weight 55% to model DC (which we regard as most independent of exogenous influences), 15%

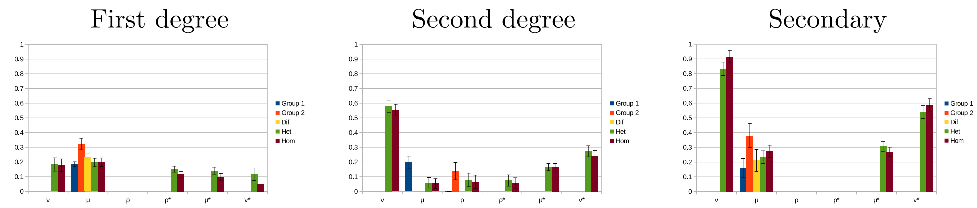

Figure 1: Comparison of regime coefficients

Parameter	First degree par.	inc.	rel.	Second degree par.	inc.	rel.	Secondary par.	inc.	rel.
ν	0.18(0.04)	0.28	0.15	0.57(0.04)	0.7	0.15	0.86(0.04)	1.41	0.15
μ	0.23(0.02)	0.36	1	0.13(0.04)	0.16	0.3	0.23(0.07)	0.38	1
ρ				0.11(0.05)	0.13	0.3			
$\rho^\star$	0.14(0.02)	0.22	0.15	0.07(0.04)	0.08	0.15			
$\mu^\star$	0.13(0.02)	0.2	0.15	0.17(0.02)	0.2	0.15	0.29(0.03)	0.48	0.15
$\nu^\star$	0.09(0.04)	0.15	0.15	0.26(0.04)	0.32	0.15	0.56(0.04)	0.91	0.15
γ	0.01(0.03)	–	–	−0.03(0.03)	–	–	−0.2(0.04)	–	–
ω	0.64(0.05)	–	–	0.85(0.07)	–	–	0.81(0.07)	–	–

Table 3: Aggregated estimates of regime coefficients.*par.* – aggregated estimate(standard error), *inc.* – percentage increase given no contact restriction (see (8)), *rel.* – reliability: the sum of source coefficients' weights.

to each of the two estimates given by D, 10% to W and 5% to WH. The results may be found in Table 3 together with the estimated percentage increase of the growth given no contact restriction (see (8)) and the reliability of the estimates, which we evaluate as the sum of weights of source estimates at our disposal.

Next, for each cohort with the exception of kindergartens. we evaluated the predictions of the growth rate ρ_t and compared them with actual values. For the predictions, we used the aggregated parameters (Table 3). The results are depicted in Figure 1. where the value ρ_t is also decomposed into individual parts, i.e. G_t and the components of S_t corresponding to individual regimes. We see that, despite the initial "shook" in the beginning of the school year, the degree the schools contributed to the growth is not high in comparison with the influence of the surrounding epidemic, especially given safer regimes including masks. Thus, it remains questionable whether the extensive school closures were reasonable. On one hand, opening would partly contribute to the epidemic growth, on the other hand, the contribution would not have been large. Moreover, since March, there was a clear "space" in the graphs suggesting that the spring opening of schools need not have been as cautious as it was.´

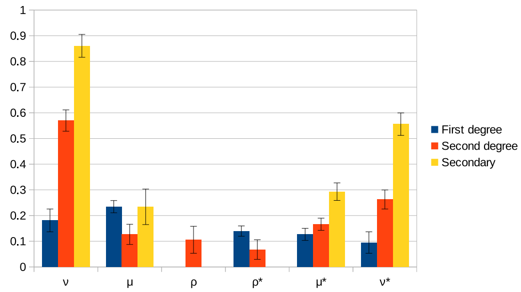

Figure 2: Aggregated estimates of regime coefficients

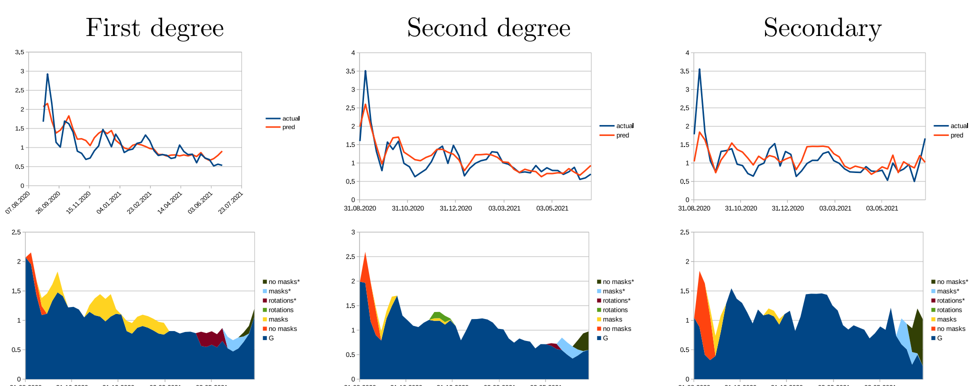

Figure 3: Growth $\frac{X_t}{X_{t-1}}$, ρ_t and its decomposition.

Discussion

The presented results chart a consistent picture of the relations between school opening and infections. Almost beyond doubt we demonstrated that opening of primary as well as secondary schools does have a significant influence on the infections within corresponding age cohorts of students. Kindergartens, on the other hand, appear to have negligible or no impact. Our results also demonstrate that the influence grows with the age of students. We are also able to quantify the influence of various regimes, especially for the case of mandatory wearing masks in classrooms, which we were able to estimate reliably by means of sub-cohort comparison.

The main limitation of our study is the possible dependence on time, hence on specific temporal circumstances. Even in the case of the most environment-independent method – evaluating differences between sub-cohorts (D and DC) – all the observations with the positive difference between covariates come from a exclusively different time period than the rest of the observations, the reason being that all the measures have been applied statewide, Consequently, the usual technique of adding time dummies thus could not be applied here as the dummies would be co-linear with the alleged difference. On the other hand, the fact that the estimates of the regime coefficients which we were able to be estimated by different models do not differ dramatically, speaks in favour of the validity of these estimates.

The results are also sensitive to the choice of the infectiousness rate d_t, which we estimated from the global history of the epidemics, regardless the age, see Appendix A. To explore this sensitivity, we re-estimated the model with a dramatically higher ascertainment rate $\alpha = 1$ (i.e. assuming that all infection were have been reported) and also (unrealistically) assuming d_t to be constant, equal to the average of our original estimate. The results can be seen in Tables 9, Table 2, respectively. We see that, yet some values changed, those, which we regard as reliable, are similar. Note also that the perturbations we used for the sensitivity analysis were extreme and hence unrealistic.

Another issue to be considered is the fact that the number of reported infections is less than the number of actual infections (ascertainment rate is less than one). This would not be problem if the rate did not change over time: If it were constant and the same for all the age cohorts, then our models, lacking a constant term and hence being linear, would hold with the same coefficients for the reported values as for the actual ones; if the rate were constant but differing between the examined cohort and the overall population, then the model would hold too, but with different coefficients. However, this issue makes a problem when the rates change over time. The problem would be less if the rates (that for the cohort and the overall one) changed gradually, keeping roughly their ratio (then there would be only the small changes between the weeks, which is what matters), but the problem would be serious if, for instance, the ratio jumped for the students but not for the rest of the population, as it is the case when testing is being introduced to schools. Then, clearly, the coefficients of the regimes including testing would not be comparable with those without testing.

Thus, in our models, we should be very cautious when comparing coefficients without stars with those with stars, which will be likely relatively overestimated. Another manifestation of this issue, very probably resulting in a change of detection ratio, was a recommendation[2] not to test children exhibiting symptoms from the beginning of September 2020; it is likely that this issue stands beyond the jumps in Figure 3 in September.

Further, there is a determinant of children infections not taken into account in our analysis: encounters with teachers, which can be significant [Brom, 2021]. In our framework this phenomenon could be modelled by adding a term $\phi A_{t-1} D_t Y_{i,t-1}$ to (3) where $A_t = N_t + \cdots + R_t^\star$ is the rate of presence at school (we take the district infections as covariate as the teachers are likely to be local). Another topic to be discussed is the hypothesis that children, not going to schools, are infected anyway in other environments. This would mean, in the language of (3), that additional $\phi D_t(1 - A_{t-1})Y_{i,t-1}$ would be infected. Both the hypotheses can be tested by adding a term $\psi D_t A_{t-1} Y_{i,t-1}$ into (3) with positive values of ψ speaking of additional infections in schools; negative values speaking for infections outside when the school is closed. Doing that, we got significant positive values for all the first degree, the second degree and secondary schools with the estimates of regime coefficients shifted. However, the fact that the new covariate is correlated with the regime ones introduced co-linearity into the estimation (confirmed by Belsley-Kuh-Welsch tests), so there is a risk that the results are spurious. Nevertheless, as the sums of ψ and the (newly estimated) regime coefficients is still significantly positive, these results speak against the hypothesis that, during school closures, children are more infected elsewhere.

Acknowledgement. The authors are grateful to Ján Palguta and Arnošt Komárek for valuable comments and suggestions.

References

[Bignardi et al., 2020] Bignardi, G., Dalmaijer, E. S., Anwyl-Irvine, A. L., Smith, T. A., Siugzdaite, R., Uh, S., and Astle, D. E. (2020). Longitudinal increases in childhood depression symptoms during the COVID-19 lockdown. Accepted: 2020-12-01T07:10:54Z Publisher: BMJ Publishing Group.

[Brauner et al., 2021a] Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., et al. (2021a). Inferring the effectiveness of government interventions against covid-19. *Science*, 371(6531).

[Brauner et al., 2021b] Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick,

[2]https://koronavirus.mzcr.cz/wp-content/uploads/2020/09/Metodick%C3%BD-postup_OSPDL_MZ_Covid-ve-%C5%A1kol%C3%A1ch.pdf?fbclid=IwAR193MQUjdD06dm9AhJTl3W60gQfJFdDJFDfpDhjX_KtJ6KAwvjOMCT1WzE

M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., and Kulveit, J. (2021b). Inferring the effectiveness of government interventions against COVID-19. *Science*, 371(6531). Publisher: American Association for the Advancement of Science Section: Research Article.

[Brom, 2021] Brom, Cyril, e. a. (2021). Rotation-based schedules in elementary schoolsto prevent covid-19 spread: A simulation study. to appear.

[Chech Ministry of Health, 2021] Chech Ministry of Health (2021). Open datasets. `https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19/osoby.csv`.

[Di Pietro et al., 2020] Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., and Mazza, J. (2020). The likely impact of covid-19 on education: Reflections based on the existing literature and recent international datasets. Technical report.

[ECDC, 2020] ECDC (2020). COVID-19 in children and the role of school settings in transmission - first update.

[Engzell et al., 2020] Engzell, P., Frey, A., and Verhagen, M. D. (2020). Learning Loss Due to School Closures During the COVID-19 Pandemic. Technical report, SocArXiv. type: article.

[Forbes et al., 2021] Forbes, H., Morton, C. E., Bacon, S., McDonald, H. I., Minassian, C., Brown, J. P., Rentsch, C. T., Mathur, R., Schultze, A., DeVito, N. J., MacKenna, B., Hulme, W. J., Croker, R., Walker, A. J., Williamson, E. J., Bates, C., Mehrkar, A., Curtis, H. J., Evans, D., Wing, K., Inglesby, P., Drysdale, H., Wong, A. Y. S., Cockburn, J., McManus, R., Parry, J., Hester, F., Harper, S., Douglas, I. J., Smeeth, L., Evans, S. J. W., Bhaskaran, K., Eggo, R. M., Goldacre, B., and Tomlinson, L. A. (2021). Association between living with children and outcomes from covid-19: OpenSAFELY cohort study of 12 million adults in England. *BMJ*, 372:n628. Publisher: British Medical Journal Publishing Group Section: Research.

[Gavenčiak et al., 2021] Gavenčiak, T., Monrad, J. T., Leech, G., Sharma, M., Mindermann, S., Brauner, J. M., Bhatt, S., and Kulveit, J. (2021). Seasonal variation in sars-cov-2 transmission in temperate climates. *medRxiv*.

[Gold, 2021] Gold, J. A. W. (2021). Clusters of SARS-CoV-2 Infection Among Elementary School Educators and Students in One School District — Georgia, December 2020–January 2021. *MMWR. Morbidity and Mortality Weekly Report*, 70.

[Haug et al., 2020] Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S., and Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, 4(12):1303–1312. Publisher: Nature Publishing Group.

[Ismail et al., 2021] Ismail, S. A., Saliba, V., Lopez Bernal, J., Ramsay, M. E., and Ladhani, S. N. (2021). SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *The Lancet Infectious Diseases*, 21(3):344–353.

[Lessler et al., 2021] Lessler, J., Grabowski, M. K., Grantz, K. H., Badillo-Goicoechea, E., Metcalf, C. J. E., Lupton-Smith, C., Azman, A. S., and Stuart, E. A. (2021). Household COVID-19 risk and in-person schooling. *medRxiv*, page 2021.02.27.21252597. Publisher: Cold Spring Harbor Laboratory Press.

[Maldonado and De Witte, 2020] Maldonado, J. and De Witte, K. (2020). The effect of school closures on standardised student test. FEB Research Report Department of Economics.

[PAQ research, 2021] PAQ research (2021). Život během pandemie. `https://zivotbehempandemie.cz/`. Sociologická studie, cit. 13.5.2021.

[Ravens-Sieberer et al., 2021] Ravens-Sieberer, U., Kaman, A., Erhart, M., Devine, J., Hölling, H., Schlack, R., Löffler, C., Hurrelmann, K., and Otto, C. (2021). Quality of life and mental health in children and adolescents during the first year of the covid-19 pandemic in germany: Results of a two-wave nationally representative study. *Preprint available at SSRN 3798710*.

[SAGE, 2020] SAGE (2020). Tfc: Children and transmission, 4 november 2020. paper prepared by the children's task and finish group (tfc) for the scientific advisory group for emergencies (sage).

[Stein-Zamir et al., 2020] Stein-Zamir, C., Abramson, N., Shoob, H., Libal, E., Bitan, M., Cardash, T., Cayam, R., and Miskin, I. (2020). A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. *Eurosurveillance*, 25(29):2001352. Publisher: European Centre for Disease Prevention and Control.

[Torres et al., 2020] Torres, J. P., Piñera, C., De La Maza, V., Lagomarcino, A. J., Simian, D., Torres, B., Urquidi, C., Valenzuela, M. T., and O'Ryan, M. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 Antibody Prevalence in Blood in a Large School Community Subject to a Coronavirus Disease 2019 Outbreak: A Cross-sectional Study. *Clinical Infectious Diseases*, (ciaa955).

[Zhu et al., 2020] Zhu, Y., Bloxham, C. J., Hulme, K. D., Sinclair, J. E., Tong, Z. W. M., Steele, L. E., Noye, E. C., Lu, J., Xia, Y., Chew, K. Y., Pickering, J., Gilks, C., Bowen, A. C., and Short, K. R. (2020). A meta-analysis on the role of children in SARS-CoV-2 in household transmission clusters. *medRxiv*. Publisher: Cold Spring Harbor Laboratory Press.

[Zimmerman et al., 2021] Zimmerman, K. O., Akinboyo, I. C., Brookhart, M. A., Boutzoukas, A. E., McGann, K. A., Smith, M. J., Maradiaga Panayotti, G., Armstrong, S. C., Bristow, H., Parker, D., Zadrozny, S., Weber,

D. J., Benjamin, D. K., and FOR THE ABC SCIENCE COLLABORATIVE (2021). Incidence and Secondary Transmission of SARS-CoV-2 Infections in Schools. *Pediatrics*, 147(4):e2020048090.

Appendix

A The Epidemic Model

To capture the influence of other infection sources than schools, we use a simple model in which the infections $X_{i,t}$ in the cohort within the district i at time t depend on the previous week overall infections, the infections within the district and the infections within the cohort in the district:

$$X_{i,t} = \alpha_i D_t Y_{t-1} + \beta_i D_t Y_{i,t-1} + \gamma D_t X_{i,t-1} + e_{i,t}, \qquad D_t = d_t C_{t-2} \quad (9)$$

Here, C_t is the overall contact reduction reported by the longitudinal sociological study [PAQ research, 2021] and d_t is the rate of infectiousness. The time lag two of the contact restriction was chosen as that maximizing the correlation: $\max_\tau \operatorname{corr}(Y_t, C_{t-\tau} Y_{t-1})$, the infection ratio was determined as $d_t = w_t r_t i_t$. Here, $w_t = (1+\varsigma\cos(at+b))$, where $\varsigma = 0.18$, a and b are set so that the period is one-year period and the peak is on January, 10th, is a cyclic component reflecting the (direct or indirect) influence of weather conditions. Further, r_t is the course of infectiousness, determined by the composition of the virus variants present in the Czech republic; which we assumed r_t to be constant, equal to $r^0 = 1.55$, up to the end of 2020, linear up to March 1st, 2021, and then constant, equal to $r^1 = 2.44$. Finally, $i_t = (1 - \alpha\iota_t)$ is the effect of natural immunization, where $\alpha = 0.4$ is the ascertainment rate and ι_t is the ratio of total reported infections within the examined cohorts (from 3 to 21 year old with only half of the latter cohort). As no children younger than 16 were vaccinated by June 2021 and only a small minority of students over 16 had got their first dose by that time, we did not include the effect of vaccination in d_t. The parameters ς, r^0, r^1 we obtained by estimation; notably, the value of ς is very close to that obtained independently by [Gavenčiak et al., 2021]. The value α has been set according the rate of respondents of [PAQ research, 2021] suffering from typical covid symptoms who underwent testing. See Figure 4 for the course of d_t and its empirical counterparts, and also Table 4 for the estimation of homogenized version of (9):

$$X_{i,t} = \alpha h s_i D_t Y_{t-1} + \beta h D_t Y_{i,t-1} + \gamma D_t X_{i,t-1} + e_{i,t} \quad (10)$$

where h is the relative size of the examined cohort with respect to the rest of the population and s_i is the relative size of the i-th district's population.

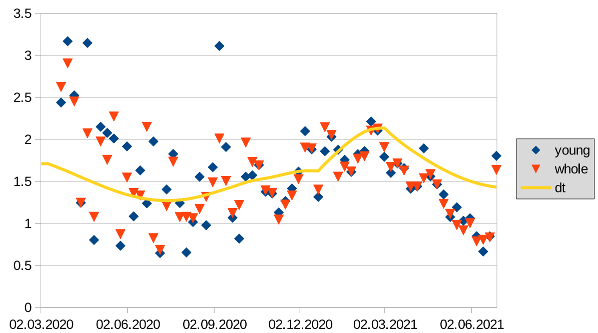

Figure 4: Course of d_t (line), $\frac{Y_t}{Y_{t-1}C_{t-2}}$ (triangles) and $\frac{\overline{X}_t}{\overline{X}_{t-1}C_{t-2}}$ (diamonds) where $\overline{X}_t = X_t^k + X_t^f + X_t^s + X_t^e$ is the number of infections in all the examined cohorts.

Parameter	Kindergartens	First degree	Second degree	Secondary
α	$0.1^{***}(0.01)$	$0.14^{***}(0.01)$	$0.19^{***}(0.02)$	$0.23^{***}(0.02)$
β	$0.37^{***}(0.01)$	$0.4^{***}(0.01)$	$0.45^{***}(0.01)$	$0.26^{***}(0.02)$
γ	$0.14^{***}(0.01)$	$0.07^{***}(0.01)$	$0.11^{***}(0.01)$	$0.23^{***}(0.01)$

Table 4: Estimation of (10)

B Detailed Inputs

In Tables 5, 6 and 7, we list covariates involved in all our analyses. In the former two, symbol $\times$ means that the observation have not been used for estimation of (1) and (3), mostly doe to the lack of knowledge of the actual situations. Meaning of abbreviations: a – voluntary, only 15 pupils in a classroom (approx. half), b – only 4 days from week, c – only 1st and 2nd classes open, d – only the last year open, e – summer vacation, f – Christmas vacation, g – autumn vacation, h – closed on Wednesday, i – masks ordered from, j – different regime among regions according to epidemiological situation, k – regime change during a week, l – last year fully open, the rest rotations, m – starting from Tuesday, valid in all but 3 regions on Monday, n – open up to decision of directors.

C Detailed Results

Table 8 shows the parameter estimates by individual models. Tables 9 and 10 show aggregated estimates in the case that d_t is computed with $\alpha = 1$, that d_t is constant, respectively.

	kindergartens		first degree						
	N		N	M	R	$R^\star$	$M^\star$	$N^\star$	
06.04.2020		$\times^n$	0	0	0	0	0	0	
13.04.2020		$\times^n$	0	0	0	0	0	0	
20.04.2020		$\times^n$	0	0	0	0	0	0	
27.04.2020		$\times^n$	0	0	0	0	0	0	
04.05.2020		$\times^n$	0	0	0	0	0	0	
11.05.2020		$\times^n$	0	0	0	0	0	0	
18.05.2020		$\times^n$	0	0	0	0	0	0	
25.05.2020		$\times^n$							$\times^a$
01.06.2020		$\times^n$							$\times^a$
08.06.2020		$\times^n$							$\times^a$
15.06.2020		$\times^n$							$\times^a$
22.06.2020		$\times^n$							$\times^a$
29.06.2020		$\times^e$							$\times^e$
06.07.2020		$\times$							$\times$
13.07.2020		$\times$							$\times$
20.07.2020		$\times$							$\times$
27.07.2020		$\times$							$\times$
03.08.2020		$\times$							$\times$
10.08.2020		$\times$							$\times$
17.08.2020		$\times$							$\times$
24.08.2020		$\times^e$							$\times^e$
31.08.2020	0.8	^{a}	0.8	0	0	0	0	0	^{b}
07.09.2020	1		1	0	0	0	0	0	
14.09.2020	1		0.6	0.4	0	0	0	0	^{I}
21.09.2020	1		0	1	0	0	0	0	
28.09.2020	1		0	0.8	0	0	0	0	^{b}
05.10.2020	1		0	1	0	0	0	0	
12.10.2020	1		0	0.2	0	0	0	0	$\times^h$
19.10.2020	1		0	0	0	0	0	0	
26.10.2020	1		0	0	0	0	0	0	$\times^g$
02.11.2020	1		0	0	0	0	0	0	
09.11.2020	1		0	0	0	0	0	0	
16.11.2020	1			0.3					$\times^k$
23.11.2020	1			0.76					$\times^k$
30.11.2020	1		0	1	0	0	0	0	
07.12.2020	1		0	1	0	0	0	0	
14.12.2020	1		0	1	0	0	0	0	
21.12.2020		$\times^f$		0.2					$\times^f$
28.12.2020		$\times^f$							$\times^f$
04.01.2021	1		0	0.4	0	0	0	0	^{c}
11.01.2021	1		0	0.4	0	0	0	0	^{c}
18.01.2021	1		0	0.4	0	0	0	0	^{c}
25.01.2021	0		0	0.4	0	0	0	0	^{c}
01.02.2021	0		0	0.4	0	0	0	0	^{c}
08.02.2021	0		0	0.4	0	0	0	0	^{c}
15.02.2021	0		0	0.4	0	0	0	0	^{c}
22.02.2021	0		0	0.4	0	0	0	0	^{c}
01.03.2021	0		0	0	0	0	0	0	
08.03.2021	0		0	0	0	0	0	0	
15.03.2021	0		0	0	0	0	0	0	
22.03.2021	0		0	0	0	0	0	0	
29.03.2021	0		0	0	0	0	0	0	
05.04.2021	0		0	0	0	0	0	0	
12.04.2021		$\times^d$	0	0	0	1	0	0	
19.04.2021		$\times^d$	0	0	0	1	0	0	
26.04.2021		$\times^j$	0	0	0	1	0	0	
03.05.2021		$\times^j$	0	0	0	1	0	0	
10.05.2021	1		0	0	0	1	0	0	
17.05.2021	1		0	0	0	0	1	0	
24.05.2021	1		0	0	0	0	1	0	
31.05.2021	1		0	0	0	0	1	0	
07.06.2021	1		0	0	0	0	0.5	0.5	$\times^{jk}$
14.06.2021	1		0	0	0	0	0.05	0.95	
21.06.2021	1		0	0	0	0	0	1	

Table 5: Covariates of (1) and (3).

	second degree							secondary						
	N	M	R	$R^\star$	$M^\star$	$N^\star$		N	M	R	$R^\star$	$M^\star$	$N^\star$	
06.04.2020	0	0	0	0	0	0		0	0	0	0	0	0	
13.04.2020	0	0	0	0	0	0		0	0	0	0	0	0	
20.04.2020	0	0	0	0	0	0		0	0	0	0	0	0	
27.04.2020	0	0	0	0	0	0		0	0	0	0	0	0	
04.05.2020	0	0	0	0	0	0		0	0	0	0	0	0	
11.05.2020							$\times^{ad}$							$\times^{ad}$
18.05.2020							$\times^{ad}$							$\times^{ad}$
25.05.2020							$\times^{ad}$							$\times^{ad}$
01.06.2020							$\times^{ad}$							$\times^{ad}$
08.06.2020							$\times^{ad}$							$\times^{e}$
15.06.2020							$\times^{ad}$							$\times^{e}$
22.06.2020							$\times^{ad}$							$\times^{e}$
29.06.2020							$\times^{e}$							$\times^{e}$
06.07.2020							$\times$							$\times$
13.07.2020							$\times$							$\times$
20.07.2020							$\times$							$\times$
27.07.2020							$\times$							$\times$
03.08.2020							$\times$							$\times$
10.08.2020							$\times$							$\times$
17.08.2020							$\times$							$\times$
24.08.2020							$\times^{e}$							$\times^{e}$
31.08.2020	0.8	0	0	0	0	0	^{b}	0.8	0	0	0	0	0	^{b}
07.09.2020	1	0	0	0	0	0		1	0	0	0	0	0	
14.09.2020	0.6	0.4	0	0	0	0	^{I}	0.6	0.4	0	0	0	0	^{I}
21.09.2020	0	1	0	0	0	0		0	1	0	0	0	0	
28.09.2020	0	0.8	0	0	0	0	^{b}	0	0.8	0	0	0	0	^{b}
05.10.2020	0	1	0	0	0	0								$\times^{j}$
12.10.2020							$\times^{h}$							$\times^{h}$
19.10.2020	0	0	0	0	0	0		0	0	0	0	0	0	
26.10.2020	0	0	0	0	0	0	$\times^{g}$	0	0	0	0	0	0	$\times^{g}$
02.11.2020	0	0	0	0	0	0		0	0	0	0	0	0	
09.11.2020	0	0	0	0	0	0		0	0	0	0	0	0	
16.11.2020	0	0	0	0	0	0		0	0	0	0	0	0	
23.11.2020	0	0	0	0	0	0								$\times^{k}$
30.11.2020	0	0.3	0.8	0	0	0	^{l}	0	0.25	0	0	0	0	^{d}
07.12.2020	0	0.3	0.8	0	0	0	^{l}	0	0.25	0	0	0	0	^{d}
14.12.2020	0	0.3	0.8	0	0	0	^{l}	0	0.25	0	0	0	0	^{d}
21.12.2020		0.1	0.2				$\times^{f}$							$\times^{f}$
28.12.2020							$\times^{f}$							$\times^{f}$
04.01.2021	0	0	0	0	0	0								$\times^{f}$
11.01.2021	0	0	0	0	0	0		0	0	0	0	0	0	
18.01.2021	0	0	0	0	0	0		0	0	0	0	0	0	
25.01.2021	0	0	0	0	0	0		0	0	0	0	0	0	
01.02.2021	0	0	0	0	0	0		0	0	0	0	0	0	
08.02.2021	0	0	0	0	0	0		0	0	0	0	0	0	
15.02.2021	0	0	0	0	0	0		0	0	0	0	0	0	
22.02.2021	0	0	0	0	0	0		0	0	0	0	0	0	
01.03.2021	0	0	0	0	0	0		0	0	0	0	0	0	
08.03.2021	0	0	0	0	0	0		0	0	0	0	0	0	
15.03.2021	0	0	0	0	0	0		0	0	0	0	0	0	
22.03.2021	0	0	0	0	0	0		0	0	0	0	0	0	
29.03.2021	0	0	0	0	0	0		0	0	0	0	0	0	
05.04.2021	0	0	0	0	0	0		0	0	0	0	0	0	
12.04.2021	0	0	0	0	0	0		0	0	0	0	0	0	
19.04.2021	0	0	0	0	0	0		0	0	0	0	0	0	
26.04.2021	0	0	0	0	0	0		0	0	0	0	0	0	
03.05.2021				0.5			$\times^{j}$	0	0	0	0	0	0	
10.05.2021	0	0	0	1	0	0		0	0	0	0	0	0	
17.05.2021	0	0	0	0	1	0		0	0	0	0	0	0	
24.05.2021	0	0	0	0	1	0		0	0	0	0	1	0	
31.05.2021	0	0	0	0	1	0		0	0	0	0	1	0	
07.06.2021	0	0	0	0	0.5	0.5	$\times^{jk}$					0.5	0.5	$\times^{jk}$
14.06.2021	0	0	0	0	0.1	1		0	0	0	0	0.05	0.95	
21.06.2021	0	0	0	0	0	1		0	0	0	0	0	1	

Table 6: Covariates of (1) and (3).

	kindergartens		first degree								
	second degree			secondary							
	N^1	N^2	M^1	M^2	ΔM	M^1	R^1	$M^1 - R^1$	M^1	M^2	ΔM
06.04.2020			0	0	0	0	0	0	0	0	0
13.04.2020			0	0	0	0	0	0	0	0	0
20.04.2020			0	0	0	0	0	0	0	0	0
27.04.2020			0	0	0	0	0	0	0	0	0
04.05.2020			0	0	0	0	0	0	0	0	0
11.05.2020			0	0	0						
18.05.2020			0	0	0						
25.05.2020											
01.06.2020											
08.06.2020											
15.06.2020											
22.06.2020											
29.06.2020											
06.07.2020											
13.07.2020											
20.07.2020											
27.07.2020											
03.08.2020											
10.08.2020											
17.08.2020											
24.08.2020											
31.08.2020	0.8	0.8			0	0	0	0	0	0	0
07.09.2020	1	1			0	0	0	0	0	0	0
14.09.2020	1	1	0.4	0.4	0	0.4	0	0	0.4	0.4	0
21.09.2020	1	1	1	1	0	1	0	0	1	1	0
28.09.2020	1	1	0.8	0.8	0	0.8	0	0	0.8	0.8	0
05.10.2020	1	1	1	1	0	1	0	0			
12.10.2020	1	1									
19.10.2020	1	1	0	0	0	0	0	0	0	0	0
26.10.2020	1	1									
02.11.2020	1	1	0	0	0	0	0	0	0	0	0
09.11.2020	1	1	0	0	0	0	0	0	0	0	0
16.11.2020						0	0	0	0	0	0
23.11.2020						0	0	0			
30.11.2020			0	0	0	1	0	1	1	0	1
07.12.2020	1	1	0	0	0	1	0	1	1	0	1
14.12.2020	1	1	0	0	0	1	0	1	1	0	1
21.12.2020											
28.12.2020											
04.01.2021	1	1	1	0	1	0	0	0			
11.01.2021	1	1	1	0	1	0	0	0	0	0	0
18.01.2021	1	1	1	0	1	0	0	0	0	0	0
25.01.2021	0	0	1	0	1	0	0	0	0	0	0
01.02.2021	0	0	1	0	1	0	0	0	0	0	0
08.02.2021	0	0	1	0	1	0	0	0	0	0	0
15.02.2021	0	0	1	0	1	0	0	0	0	0	0
22.02.2021	0	0	1	0	1	0	0	0	0	0	0
01.03.2021	0	0	0	0	0	0	0	0	0	0	0
08.03.2021	0	0	0	0	0	0	0	0	0	0	0
15.03.2021	0	0	0	0	0	0	0	0	0	0	0
22.03.2021	0	0	0	0	0	0	0	0	0	0	0
29.03.2021	0	0	0	0	0	0	0	0	0	0	0
05.04.2021	0	0	0	0	0	0	0	0	0	0	0
12.04.2021	1	0			0	0	0	0	0	0	0
19.04.2021	1	0			0	0	0	0	0	0	0
26.04.2021					0	0	0	0	0	0	0
03.05.2021					0				0	0	0
10.05.2021	1	1			0	0	0	0	0	0	0
17.05.2021	1	1			0			0	0	0	0
24.05.2021	1	1			0			0			0
31.05.2021	1	1			0			0			0
07.06.2021	1	1									
14.06.2021	1	1			0			0			0
21.06.2021	1	1			0			0			0

Table 7: Covariates of the sub-cohort analyses.

Model		Kindergartens	First degree	Second degree	Secondary
S	ν^1	$0.01(0.04)$			
($w = 15\%$	ν^2	$-0.01(0.04)$			
Each)	μ^1		$0.18^{***}(0.02)$	$0.2^{***}(0.04)$	$0.16^{**}(0.06)$
	μ^2		$0.32^{***}(0.04)$		$0.38^{***}(0.08)$
	ρ^2			$0.14^{**}(0.06)$	
C	μ		$0.23^{***}(0.02)$		$0.21^{***}(0.07)$
($w = 55\%$)	$\mu - \rho$			$0.04(0.06)$	
W	ν	$0(0)$	$0.18^{***}(0.04)$	$0.58^{***}(0.04)$	$0.83^{***}(0.05)$
($w = 10\%$)	μ		$0.2^{***}(0.03)$	$0.06(0.04)$	$0.23^{***}(0.04)$
	ρ			$0.08^{*}(0.05)$	
	$\rho^\star$		$0.15^{***}(0.02)$	$0.07^{*}(0.04)$	
	$\mu^\star$		$0.14^{***}(0.02)$	$0.17^{***}(0.02)$	$0.31^{***}(0.03)$
	$\nu^\star$		$0.12^{***}(0.04)$	$0.27^{***}(0.04)$	$0.54^{***}(0.04)$
	γ	$0.15^{***}(0.02)$	$-0.12^{***}(0.03)$	$-0.1^{***}(0.03)$	$-0.24^{***}(0.04)$
WH	ν	$-0.01^{**}(0)$	$0.18^{***}(0.04)$	$0.55^{***}(0.04)$	$0.92^{***}(0.04)$
($w = 5\%$)	μ		$0.2^{***}(0.03)$	$0.05(0.03)$	$0.27^{***}(0.04)$
	ρ			$0.06(0.05)$	
	$\rho^\star$		$0.12^{***}(0.02)$	$0.05(0.04)$	
	$\mu^\star$		$0.1^{***}(0.02)$	$0.17^{***}(0.02)$	$0.27^{***}(0.03)$
	$\nu^\star$		$0.05(0.04)$	$0.24^{***}(0.04)$	$0.59^{***}(0.04)$
	α	$0.02(0.01)$	$0.09^{***}(0.02)$	$0.12^{***}(0.03)$	$0.25^{***}(0.03)$
	β	$0.15^{***}(0.01)$	$0.54^{***}(0.03)$	$0.74^{***}(0.04)$	$0.56^{***}(0.04)$
	γ	$0.24^{***}(0.02)$	$0.01(0.03)$	$-0.03(0.03)$	$-0.2^{***}(0.04)$

Table 8: Details of parameter estimation

Para-meter	First degree			Second degree			Secondary		
	par.	inc.	rel.	par.	inc.	rel.	par.	inc.	rel.
ν	0.28(0.04)	0.5	0.15	0.64(0.04)	0.9	0.15	0.89(0.04)	1.71	0.15
μ	0.22(0.02)	0.4	1	0.16(0.04)	0.22	0.3	0.23(0.06)	0.45	1
ρ				0.11(0.05)	0.16	0.3			
$\rho^\star$	0.11(0.02)	0.19	0.15	0.04(0.03)	0.06	0.15			
$\mu^\star$	0.11(0.02)	0.2	0.15	0.14(0.02)	0.19	0.15	0.23(0.03)	0.45	0.15
$\nu^\star$	0.09(0.03)	0.16	0.15	0.23(0.03)	0.32	0.15	0.44(0.04)	0.85	0.15
γ	−0.03(0.03)	–	–	−0.07(0.03)	–	–	−0.19(0.04)	–	–
ω	0.58(0.04)	–	–	0.79(0.06)	–	–	0.71(0.07)	–	–

Table 9: Sensitivity analysis: $\alpha = 1$ (all cases reported)

Para-meter	First degree			Second degree			Secondary		
	par.	inc.	rel.	par.	inc.	rel.	par.	inc.	rel.
ν	−0.04(0.04)	−0.05	0.15	0.32(0.04)	0.34	0.15	0.65(0.04)	0.92	0.15
μ	0.27(0.03)	0.36	1	0.04(0.04)	0.04	0.3	0.21(0.07)	0.3	1
ρ				0.08(0.05)	0.09	0.3			
$\rho^\star$	0.09(0.02)	0.12	0.15	−0.01(0.04)	−0.01	0.15			
$\mu^\star$	0.03(0.02)	0.04	0.15	0.05(0.02)	0.06	0.15	0.22(0.03)	0.31	0.15
$\nu^\star$	−0.02(0.04)	−0.03	0.15	0.12(0.04)	0.13	0.15	0.44(0.04)	0.63	0.15
γ	0.12(0.03)	–	–	0.1(0.03)	–	–	−0.13(0.04)	–	–
ω	0.62(0.05)	–	–	0.84(0.07)	–	–	0.84(0.07)	–	–

Table 10: Sensitivity analysis: d_t is constant